JOURNAL OF COMBINATORIAL THEORY 8, 99–103 (1970)

The Distance between Points in Random Trees*

A. MEIR AND J. W. MOON

University of Alberta, Edmonton, Alberta, Canada

Received October 30, 1968

ABSTRACT

Let γ denote the number of points in the path joining two arbitrary points in a random tree T_n with n labeled points. It is shown, among other things, that $E(\gamma) \sim (\frac{1}{2}\pi n)^{1/2}$.

1. INTRODUCTION

A tree is a connected graph that has no cycles. If u and v are any two points in a tree T_n with n labeled points, there is a unique path in T_n joining u and v. Let $\gamma(T_n; u, v)$ denote the number of points in this path; the distance $d(T_n; u, v)$ between u and v is the number of edges in this path, that is,

$$d(T_n; u, v) = \gamma(T_n; u, v) - 1.$$

Our object is to investigate the distribution of $\gamma(T_n; u, v)$—or, equivalently, of $d(T_n; u, v)$—for any two fixed points u and v over the set of the n^{n-2} labeled trees T_n. In particular, we will show that the average value of $\gamma(T_n; u, v)$ is asymptotically equal to $(\frac{1}{2}\pi n)^{1/2}$.

Rényi and Szekeres [5] have considered the more difficult problem of determining the distribution of

$$h(T_n; u) = \max_{v \in T_n} d(T_n; u, v),$$

the height of T_n with respect to the point u. They showed, among other things, that the average value of $h(T_n; u)$ is asymptotically equal to $(2\pi n)^{1/2}$. The problem of determining the distribution of

$$D(T_n) = \max_{u \in T_n} h(T_n; u),$$

the diameter of T_n, appears to be still unsolved.

* Presented at the Yale University Conference on Combinatorial Theory in honor of Professor Oystein Ore (May, 1968).

2. The Probability That $\gamma(T_n\,; u, v) = k$

The following lemma is proved in [3] (see also [2] and [6]):

LEMMA. *If the forest F_n consists of l trees with $j_1, j_2, \ldots, j_l$ labeled points, respectively, where $j_1 + j_2 + \cdots + j_l = n$, then there are $j_1 j_2 \cdots j_l n^{l-2}$ trees T_n that contain F_n.*

For notational convenience we let $(x)_r = x(x-1)\cdots(x-r+1)$ for positive integers r.

THEOREM 1. *If $p(n, k)$ denotes the probability that $\gamma(T_n\,; u, v) = k$, then*

$$p(n, k) = \frac{k}{n-1} \cdot \frac{(n)_k}{n^k}, \qquad \text{for} \quad k = 2, 3, \ldots, n.$$

PROOF: There are $(n-2)_{k-2}$ ways to construct a path from u to v that passes through $k-2$ of the remaining $n-2$ points. It follows from the lemma that there are kn^{n-k-1} trees T_n that contain any given path on k points. Hence, there are

$$(n-2)_{k-2}\, kn^{n-k-1} = kn^{n-k-2}(n)_k/(n-1)$$

trees T_n for which $\gamma(T_n\,; u, v) = k$. When we divide this last expression by n^{n-2}, the total number of trees T_n, we obtain the given formula for $p(n, k)$.

It follows from Theorem 1 that $p(n, 2) = 2/n$ and

$$p(n, k+1) = \frac{k+1}{k} \cdot \frac{n-k}{n}\, p(n, k), \qquad \text{for} \quad k = 2, 3, \ldots, n-1;$$

these relations are useful for computational purposes.

3. The Distribution of $\gamma(T_n\,; u, v)$

Since $e^{-t/(1-t)} < 1 - t < e^{-t}$ when $0 < t < 1$, it follows from Theorem 1 that

$$\frac{k}{n-1}\, e^{-k^2/2(n-k)} < p(n, k) < \frac{k}{n-1}\, e^{-\frac{1}{n}\binom{k}{2}}, \qquad \text{for} \quad k = 2, 3, \ldots, n.$$

Let $\alpha_n = n^{\frac{1}{2}-\epsilon}$ and $\beta_n = n^{\frac{1}{2}+\epsilon}$, where $0 < \epsilon < 1/20$. If $k \geqslant \beta_n + 1$, then

$$p(n, k) < \frac{n}{n-1}\, e^{-\frac{1}{n}\binom{k}{2}} \leqslant 2e^{-\beta_n^2/2n} = 2e^{-\frac{1}{2}n^{2\epsilon}}. \tag{1}$$

Thus,

$$\sum_{k \geqslant \beta_n + 1} kp(n, k) < 2n^2 e^{-\frac{1}{2}n^{2\epsilon}} = \sigma(1), \qquad \text{as} \quad n \to \infty, \tag{2}$$

and

$$\sum_{k \geqslant \beta_n + 1} k^2 p(n, k) < 2n^3 e^{-\frac{1}{2}n^{2\epsilon}} = \sigma(1), \qquad \text{as} \quad n \to \infty. \tag{3}$$

If $k \leqslant \alpha_n$, then $p(n, k) < k/(n-1)$. Thus

$$\sum_{k \leqslant \alpha_n} kp(n, k) < \alpha_n^3/(n-1) \leqslant 2n^{\frac{1}{2}-3\epsilon} = \sigma(n^{\frac{1}{2}}), \qquad \text{as} \quad n \to \infty, \tag{4}$$

and

$$\sum_{k \leqslant \alpha_n} k^2 p(n, k) < \alpha_n^4/(n-1) \leqslant 2n^{1-4\epsilon} = \sigma(n), \qquad \text{as} \quad n \to \infty. \tag{5}$$

If $\alpha_n < k < \beta_n + 1$, then

$$\frac{k}{n} e^{-k^2/2n}\{1 + 0(n^{-\frac{1}{2}+3\epsilon})\} = \frac{k}{n} e^{-k^2/2(n-k)} < p(n, k)$$
$$< \frac{k}{n-1} e^{-\frac{1}{n}\binom{k}{2}} = \frac{k}{n} e^{-k^2/2n}\{1 + 0(n^{-\frac{1}{2}+\epsilon})\}.$$

Therefore,

$$p(n, k) = \frac{k}{n} e^{-k^2/2n} + 0(n^{-1+4\epsilon}), \qquad \text{if} \quad \alpha_n < k < \beta_n + 1. \tag{6}$$

Thus,

$$n^{-\frac{1}{2}} \sum{}' kp(n, k) = n^{-\frac{1}{2}} \sum{}' \frac{k^2}{n} e^{-k^2/2n} + 0(n^{-\frac{1}{2}+6\epsilon}) \tag{7}$$

and

$$n^{-1} \sum{}' k^2 p(n, k) = n^{-1} \sum{}' \frac{k^3}{n} e^{-k^2/2n} + 0(n^{-\frac{1}{2}+7\epsilon}), \tag{8}$$

where the sums are over integers k such that $\alpha_n < k < \beta_n + 1$. Now

$$\left| \int_{kn^{-1/2}}^{(k+1)n^{-1/2}} u^2 e^{-u^2/2}\, du - k^2 n^{-3/2} e^{-k^2/2n} \right| \leqslant Cn^{-1},$$

for some absolute constant C. It follows, therefore, from (7) that

$$\begin{aligned} n^{-\frac{1}{2}} \sum_{\alpha_n < k < \beta_n + 1} kp(n, k) &= \int_{\alpha_n n^{-1/2}}^{(\beta_n+1)n^{-1/2}} u^2 e^{-u^2/2}\, du + 0(n^{-\frac{1}{2}+6\epsilon}) \\ &= \int_0^\infty u^2 e^{-u^2/2}\, du + \sigma(1), \end{aligned} \tag{9}$$

and, similarly, from (8), that

$$n^{-1} \sum_{\alpha_n < k < \beta_n + 1} k^2 p(n, k) = \int_0^\infty u^3 e^{-u^2/2}\, du + o(1). \tag{10}$$

THEOREM 2. The mean and variance of $\gamma = \gamma(T_n\,; u, v)$ satisfy the relations

$$E(\gamma) = \{1 + o(1)\}(\tfrac{1}{2}\pi n)^{1/2}$$

and

$$\sigma^2(\gamma) = \{1 + o(1)\}(2 - \tfrac{1}{2}\pi)\, n, \qquad \text{as } n \to \infty.$$

This result follows from inequalities (2), (3), (4), (5), (9), and (10). The following result may be proved in essentially the same way:

THEOREM 3. *If $P(n, x)$ denotes the probability that $\gamma(T_n\,; u, v) \leqslant xn^{1/2}$, then*

$$P(n, x) = 1 - e^{-x^2/2} + o(1), \qquad \text{as} \quad n \to \infty,$$

for any constant x.

Notice that it follows from equation (6) and the remark at the end of Section 2 that

$$\max_k p(n, k) = (1 + o(1))(en)^{-\frac{1}{2}};$$

if $t = [n^{\frac{1}{2}}]$, the maximum occurs when $k = t$ or $t + 1$ according as $t(t + 1) \geqslant n$ or $t(t + 1) \leqslant n$.

4. VALUES OF $E(\gamma)$ AND $E(\gamma^2)$

The entries in the accompanying table were calculated by Mr. J. Hubert.

n	$E(\gamma)$	$E(\gamma^2)$	n	$E(\gamma)$	$E(\gamma^2)$
2	2	4	12	4.312	21.688
3	2.333	5.667	14	4.642	25.358
4	2.625	7.375	16	4.951	29.049
5	2.888	9.112	18	5.243	32.757
6	3.130	10.870	20	5.520	36.480
7	3.354	12.646	25	6.159	45.841
8	3.566	14.434	50	8.697	93.303
9	3.766	16.234	100	12.323	189.677
10	3.956	18.044	150	15.119	286.881

5. Concluding Remarks

Let $\lambda = \lambda_k(T_n)$ denote the number of paths of length $k - 1$ in a random tree T_n, for $k = 2, 3, \ldots, n$. It can be shown, by arguments similar to those used in the proof of Theorem 1, that

$$E(\lambda) = \frac{1}{2} kn \frac{(n)_k}{n^k},$$

and that

$$\sigma^2(\lambda) = \frac{n}{24} k(k-1)^2 (k-2) + 0(1)$$

for fixed values of k as $n \to \infty$.

We remark in closing that the arguments used here are related to arguments Rényi [4] used in studying the distribution of the length of the cycle in connected graphs that have a unique cycle (see also Katz [1]).

References

1. L. Katz, Probability of Indecomposability of a Random Mapping Function, *Ann. Math. Statist.* **26** (1955), 512–517.
2. J. W. Moon, The Second Moment of the Complexity of a Graph, *Mathematika* **11** (1964), 95–98.
3. J. W. Moon, Enumerating Labelled Trees, *Graph Theory and Theoretical Physics* (F. Harary, ed.), Academic Press, London/New York, 1967, pp. 261–272.
4. A. Rényi, On Connected Graphs, I, *Publ. Math. Inst. Hungar. Acad. Sci.* **4** (1959), 385–388.
5. A. Rényi and G. Szekeres, On the Height of Trees, *J. Australian Math. Soc.* **7** (1967), 497–507.
6. J. Sedláček, O kostrách konečných grafů, *Časopis Pěst. Mat.* **92** (1967), 338–342.